

# Supplementary Material for: Dessie: Disentanglement for Articulated 3D Horse Shape and Pose Estimation from Images

Ci Li<sup>1</sup>, Yi Yang<sup>1,2</sup>, Zehang Weng<sup>1</sup>,  
Elin Hernlund<sup>3</sup>, Silvia Zuffi<sup>4</sup>, and Hedvig Kjellström<sup>1,3</sup>

<sup>1</sup> KTH, Sweden {cil, yiya, zehang, hedvig}@kth.se

<sup>2</sup> Scania, Sweden carol-yi.yang@scania.com

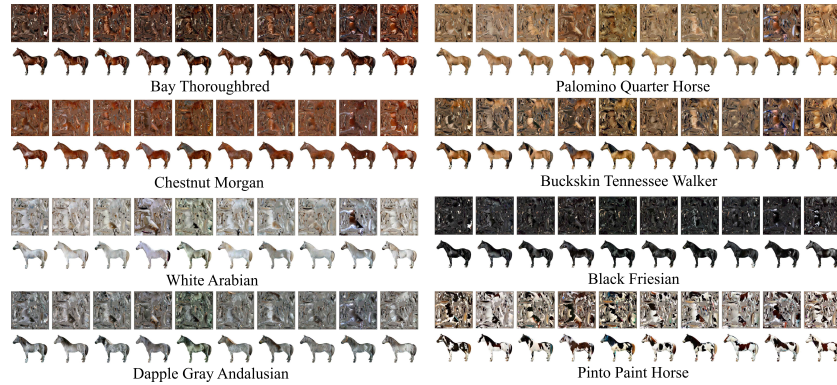
<sup>3</sup> SLU, Sweden Elin.Hernlund@slu.se

<sup>4</sup> IMATI-CNR, Italy silvia@mi.imati.cnr.it

This document provides the supplementary materials. Section 1 presents how we employ a diffusion model in DessiePIPE to synthesize realistic UV texture maps. In Section 2, we offer additional qualitative results using out-of-domain images. Section 3 offers visualization of the DINO key features for DinoHMR. Section 4 offers additional analysis regarding training Dessie with 3D GT label.

## 1 Texture Generation

We utilized the TEXTure [3] to create 80 realistic UV texture maps for our DessiePIPE. We construct specific prompts for eight unique horse species by employing the format: *A photo of a <SPECIES NAME>, {} view*.



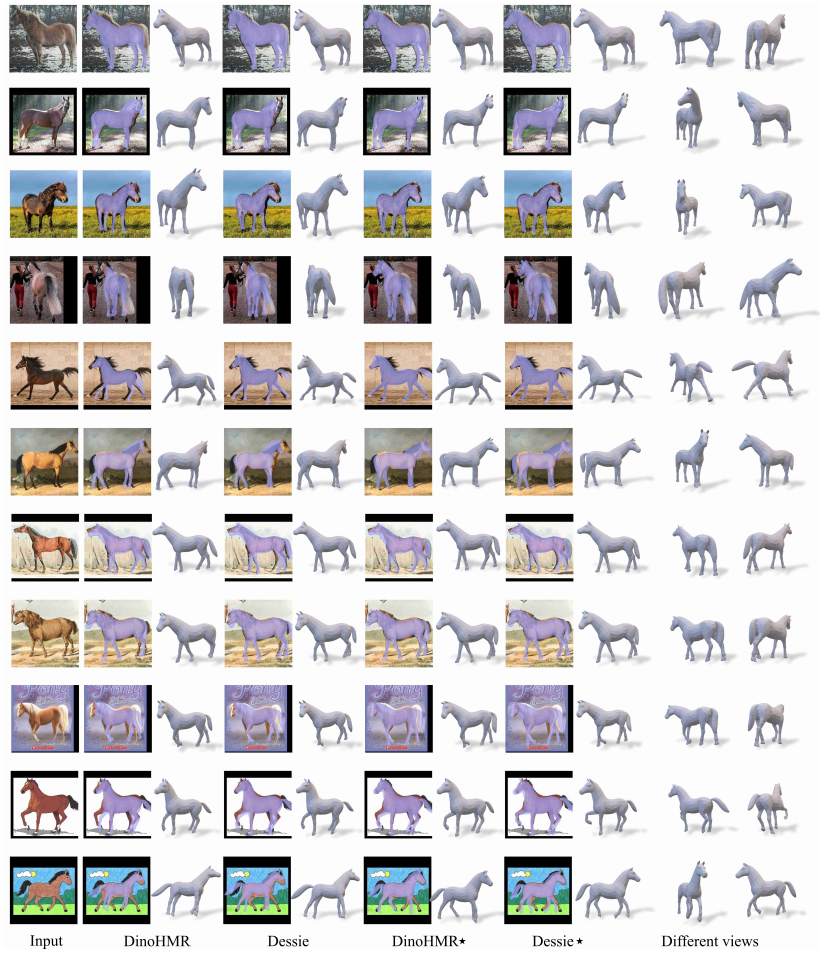
**Fig. 1:** UV texture maps created using TEXTure with eight horse species and rendered with the hSMAL model in zero pose.

<SPECIES NAME> is substituted with specific names, including "Bay Thoroughbred", "Palomino Quarter Horse", "Chestnut Morgan", "Buckskin Tennessee Walker", "White Arabian", "Black Friesian", "Dapple Gray Andalusian", and "Pinto Paint Horse". For each prompt, *{}view* is filled in with "front", "left",

"back", "right", "overhead" and "bottom", enabling the creation of textures for the horse model from various viewpoints through rotation. For each species, ten texture maps are generated using ten different random seeds. The resulting texture maps are shown in Fig. 1.

## 2 Qualitative Results

We show more reconstruction results of images from different domains in Fig. 2.



**Fig. 2:** More qualitative results of DinoHMR and Dessie before and after fine-tuning with real-world data. The first four rows showcase results from real-world images, while the subsequent rows are for images of horses from various domains, such as oil paintings and cartoons.

